



IFWO

RAW SEQUENCE LISTING

DATE: 08/23/2004

PATENT APPLICATION: US/10/656,029

TIME: 10:36:28

Input Set : A:\PTO.YF.txt

Output Set: N:\CRF4\08232004\J656029.raw

3 <110> APPLICANT: VERTEX PHARMACEUTICALS INC.

5 <120> TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR RAPID DEVELOPMENT OF
6 SCREENING ASSAYS

8 <130> FILE REFERENCE: VPI/02-143WO2

C--> 10 <140> CURRENT APPLICATION NUMBER: US/10/656,029

11 <141> CURRENT FILING DATE: 2003-09-05

13 <150> PRIOR APPLICATION NUMBER: 60/408,297

14 <151> PRIOR FILING DATE: 2002-09-05

16 <160> NUMBER OF SEQ ID NOS: 86

18 <170> SOFTWARE: PatentIn version 3.2

20 <210> SEQ ID NO: 1

21 <211> LENGTH: 795

22 <212> TYPE: DNA

23 <213> ORGANISM: Escherichia coli

25 <220> FEATURE:

26 <221> NAME/KEY: CDS

27 <222> LOCATION: (1)..(795)

28 <223> OTHER INFORMATION: B-lactamase

30 <400> SEQUENCE: 1

31 atg agt cac cca gaa acg ctg gtg aaa gta aaa gat gct gaa gat cag 48

32 Met Ser His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln

33 1 5 10 15

35 ttg ggt gca cga gtg ggt tac atc gaa ctg gat ctc aac agc ggt aag 96

36 Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys

37 20 25 30

39 atc ctt gag agt ttt cgc ccc gaa gaa cgt ttt cca atg atg agc act 144

40 Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr

41 35 40 45

43 ttt aaa gtt ctg cta tgt ggc gcg gta tta tcc cgt gtt gac gcc ggg 192

44 Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser Arg Val Asp Ala Gly

45 50 55 60

47 caa gag caa ctc ggt cgc cgc ata cac tat tct cag aat gac ttg gtt 240

48 Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val

49 65 70 75 80

51 gag tac tca cca gtc aca gaa aag cat ctt acg gat ggc atg aca gta 288

52 Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr Asp Gly Met Thr Val

53 85 90 95

55 aga gaa tta tgc agt gct gcc ata acc atg agt gat aac act gcg gcc 336

56 Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser Asp Asn Thr Ala Ala

57 100 105 110

59 aac tta ctt ctg aca acg atc gga gga ccg aag gag cta acc gct ttt 384

60 Asn Leu Leu Leu Thr Thr Ile Gly Gly Pro Lys Glu Leu Thr Ala Phe

61 115 120 125

(pg. 6)

ENTERED

RAW SEQUENCE LISTING

DATE: 08/23/2004

PATENT APPLICATION: US/10/656,029

TIME: 10:36:28

Input Set : A:\PTO.YF.txt

Output Set: N:\CRF4\08232004\J656029.raw

```

63 ttg cac aac atg ggg gat cat gta act cgc ctt gat cgt tgg gaa ccg      432
64 Leu His Asn Met Gly Asp His Val Thr Arg Leu Asp Arg Trp Glu Pro
65      130      135      140
67 gag ctg aat gaa gcc ata cca aac gac gag cgt gac acc acg atg cct      480
68 Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg Asp Thr Thr Met Pro
69 145      150      155      160
71 gca gca atg gca aca acg ttg cgc aaa cta tta act ggc gaa cta ctt      528
72 Ala Ala Met Ala Thr Thr Leu Arg Lys Leu Leu Thr Gly Glu Leu Leu
73      165      170      175
75 act cta gct tcc cgg caa caa tta ata gac tgg atg gag gcg gat aaa      576
76 Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp Met Glu Ala Asp Lys
77      180      185      190
79 gtt gca gga cca ctt ctg cgc tcg gcc ctt ccg gct ggc tgg ttt att      624
80 Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly Trp Phe Ile
81      195      200      205
83 gct gat aaa tct gga gcc ggt gag cgt ggg tct cgc ggt atc att gca      672
84 Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly Ile Ile Ala
85      210      215      220
87 gca ctg ggg cca gat ggt aag ccc tcc cgt atc gta gtt atc tac acg      720
88 Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile Tyr Thr
89 225      230      235      240
91 acg ggg agt cag gca act atg gat gaa cga aat aga cag atc gct gag      768
92 Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala Glu
93      245      250      255
95 ata ggt gcc tca ctg att aag cat tgg      795
96 Ile Gly Ala Ser Leu Ile Lys His Trp
97      260      265
100 <210> SEQ ID NO: 2
101 <211> LENGTH: 265
102 <212> TYPE: PRT
103 <213> ORGANISM: Escherichia coli
105 <400> SEQUENCE: 2
106 Met Ser His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln
107 1      5      10      15
109 Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys
110      20      25      30
112 Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr
113      35      40      45
115 Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser Arg Val Asp Ala Gly
116      50      55      60
118 Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val
119 65      70      75      80
121 Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr Asp Gly Met Thr Val
122      85      90      95
124 Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser Asp Asn Thr Ala Ala
125      100      105      110
127 Asn Leu Leu Leu Thr Thr Ile Gly Gly Pro Lys Glu Leu Thr Ala Phe
128      115      120      125
130 Leu His Asn Met Gly Asp His Val Thr Arg Leu Asp Arg Trp Glu Pro

```

RAW SEQUENCE LISTING

DATE: 08/23/2004

PATENT APPLICATION: US/10/656,029

TIME: 10:36:28

Input Set : A:\PTO.YF.txt

Output Set : N:\CRF4\08232004\J656029.raw

```

131      130      135      140
133 Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg Asp Thr Thr Met Pro
134 145      150      155      160
136 Ala Ala Met Ala Thr Leu Arg Lys Leu Leu Thr Gly Glu Leu Leu
137      165      170      175
139 Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp Met Glu Ala Asp Lys
140      180      185      190
142 Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly Trp Phe Ile
143      195      200      205
145 Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly Ile Ile Ala
146      210      215      220
148 Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile Tyr Thr
149 225      230      235      240
151 Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala Glu
152      245      250      255
154 Ile Gly Ala Ser Leu Ile Lys His Trp
155      260      265
158 <210> SEQ ID NO: 3
159 <211> LENGTH: 858
160 <212> TYPE: DNA
161 <213> ORGANISM: Escherichia coli
163 <220> FEATURE:
164 <221> NAME/KEY: CDS
165 <222> LOCATION: (1)..(858)
166 <223> OTHER INFORMATION: B-lactamase
168 <400> SEQUENCE: 3
169 atg aga att caa cat ttc cgt gtc gcc ctt att ccc ttt ttt gcg gca      48
170 Met Arg Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala
171 1      5      10      15
173 ttt tgc ctt cct gtt ttt ggt cac cca gaa acg ctg gtg aaa gta aaa      96
174 Phe Cys Leu Pro Val Phe Gly His Pro Glu Thr Leu Val Lys Val Lys
175      20      25      30
177 gat gct gaa gat cag ttg ggt gca cga gtg ggt tac atc gaa ctg gat      144
178 Asp Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp
179      35      40      45
181 ctc aac agc ggt aag atc ctt gag agt ttt cgc ccc gaa gaa cgt ttt      192
182 Leu Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe
183      50      55      60
185 cca atg atg agc act ttt aaa gtt ctg cta tgt ggc gcg gta tta tcc      240
186 Pro Met Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser
187 65      70      75      80
189 cgt gtt gac gcc ggg caa gag caa ctc ggt cgc cgc ata cac tat tct      288
190 Arg Val Asp Ala Gly Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser
191      85      90      95
193 cag aat gac ttg gtt gag tac tca cca gtc aca gaa aag cat ctt acg      336
194 Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr
195      100      105      110
197 gat ggc atg aca gta aga gaa tta tgc agt gct gcc ata acc atg agt      384
198 Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser

```

RAW SEQUENCE LISTING

DATE: 08/23/2004

PATENT APPLICATION: US/10/656,029

TIME: 10:36:28

Input Set : A:\PTO.YF.txt

Output Set: N:\CRF4\08232004\J656029.raw

```

199          115          120          125
201 gat aac act gcg gcc aac tta ctt ctg aca acg atc gga gga ccg aag      432
202 Asp Asn Thr Ala Ala Asn Leu Leu Thr Thr Ile Gly Gly Pro Lys
203          130          135          140
205 gag cta acc gct ttt ttg cac aac atg ggg gat cat gta act cgc ctt      480
206 Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr Arg Leu
207 145          150          155          160
209 gat cgt tgg gaa ccg gag ctg aat gaa gcc ata cca aac gac gag cgt      528
210 Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg
211          165          170          175
213 gac acc acg atg cct gca gca atg gca aca acg ttg cgc aaa cta tta      576
214 Asp Thr Thr Met Pro Ala Ala Met Ala Thr Thr Leu Arg Lys Leu Leu
215          180          185          190
217 act ggc gaa cta ctt act cta gct tcc cgg caa caa tta ata gac tgg      624
218 Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp
219          195          200          205
221 atg gag gcg gat aaa gtt gca gga cca ctt ctg cgc tcg gcc ctt ccg      672
222 Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro
223          210          215          220
225 gct ggc tgg ttt att gct gat aaa tct gga gcc ggt gag cgt ggg tct      720
226 Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser
227 225          230          235          240
229 cgc ggt atc att gca gca ctg ggg cca gat ggt aag ccc tcc cgt atc      768
230 Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile
231          245          250          255
233 gta gtt atc tac acg acg ggg agt cag gca act atg gat gaa cga aat      816
234 Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn
235          260          265          270
237 aga cag atc gct gag ata ggt gcc tca ctg att aag cat tgg      858
238 Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp
239          275          280          285
242 <210> SEQ ID NO: 4
243 <211> LENGTH: 286
244 <212> TYPE: PRT
245 <213> ORGANISM: Escherichia coli
247 <400> SEQUENCE: 4
248 Met Arg Ile Gln His Phe Arg Val Ala Leu Ile Pro Phe Phe Ala Ala
249 1          5          10          15
251 Phe Cys Leu Pro Val Phe Gly His Pro Glu Thr Leu Val Lys Val Lys
252          20          25          30
254 Asp Ala Glu Asp Gln Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp
255          35          40          45
257 Leu Asn Ser Gly Lys Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe
258          50          55          60
260 Pro Met Met Ser Thr Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser
261 65          70          75          80
263 Arg Val Asp Ala Gly Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser
264          85          90          95
266 Gln Asn Asp Leu Val Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr

```

RAW SEQUENCE LISTING

DATE: 08/23/2004

PATENT APPLICATION: US/10/656,029

TIME: 10:36:28

Input Set : A:\PTO.YF.txt

Output Set: N:\CRF4\08232004\J656029.raw

```

267          100          105          110
269 Asp Gly Met Thr Val Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser
270          115          120          125
272 Asp Asn Thr Ala Ala Asn Leu Leu Thr Thr Ile Gly Gly Pro Lys
273          130          135          140
275 Glu Leu Thr Ala Phe Leu His Asn Met Gly Asp His Val Thr Arg Leu
276 145          150          155          160
278 Asp Arg Trp Glu Pro Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg
279          165          170          175
281 Asp Thr Thr Met Pro Ala Ala Met Ala Thr Thr Leu Arg Lys Leu Leu
282          180          185          190
284 Thr Gly Glu Leu Leu Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp
285          195          200          205
287 Met Glu Ala Asp Lys Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro
288          210          215          220
290 Ala Gly Trp Phe Ile Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser
291 225          230          235          240
293 Arg Gly Ile Ile Ala Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile
294          245          250          255
296 Val Val Ile Tyr Thr Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn
297          260          265          270
299 Arg Gln Ile Ala Glu Ile Gly Ala Ser Leu Ile Lys His Trp
300          275          280          285
303 <210> SEQ ID NO: 5
304 <211> LENGTH: 795
305 <212> TYPE: DNA
306 <213> ORGANISM: Escherichia coli
308 <220> FEATURE:
309 <221> NAME/KEY: CDS
310 <222> LOCATION: (1)..(795)
311 <223> OTHER INFORMATION: B-lactamase
313 <400> SEQUENCE: 5
314 atg ggg cac cca gaa acg ctg gtg aaa gta aaa gat gct gaa gat cag      48
315 Met Gly His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln
316 1          5          10          15
318 ttg ggt gca cga gtg ggt tac atc gaa ctg gat ctc aac agc ggt aag      96
319 Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys
320          20          25          30
322 atc ctt gag agt ttt cgc ccc gaa gaa cgt ttt cca atg atg agc act      144
323 Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr
324          35          40          45
326 ttt aaa gtt ctg cta tgt ggc gcg gta tta tcc cgt gat gac gcc ggg      192
327 Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser Arg Asp Asp Ala Gly
328          50          55          60
330 caa gag caa ctc ggt cgc cgc ata cac tat tct cag aat gac ttg gtt      240
331 Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val
332 65          70          75          80
334 gag tac tca cca gtc aca gaa aag cat ctt acg gat ggc atg aca gta      288
335 Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr Asp Gly Met Thr Val

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/656,029

DATE: 08/23/2004
TIME: 10:36:29

Input Set : A:\PTO.YF.txt
Output Set: N:\CRF4\08232004\J656029.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:27; N Pos. 4,5,6,7,9,10,11,12,14

Seq#:28; N Pos. 6,7,8,9

Seq#:39; N Pos. 892

Seq#:41; N Pos. 9040

Seq#:51; N Pos. 2626

Seq#:73; N Pos. 1528

Seq#:76; N Pos. 8885

Seq#:79; N Pos. 9953

VERIFICATION SUMMARY

DATE: 08/23/2004

PATENT APPLICATION: US/10/656,029

TIME: 10:36:29

Input Set : A:\PTO.YF.txt

Output Set: N:\CRF4\08232004\J656029.raw

L:10 M:270 C: Current Application Number differs, Replaced Current Application Number
L:1874 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:27 after pos.:0
L:1892 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:28 after pos.:0
L:2044 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:39 after pos.:840
L:2421 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:41 after pos.:9000
L:3393 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:51 after pos.:2580
L:4264 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:73 after pos.:1500
L:4790 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:76 after pos.:8880
L:5239 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:79 after pos.:9900